Todo：

0. Read d2l transformer

1. Preprocess dataset(text cleaning, image reshaping)

2. Set up FID score evaluation, compare with and without text summarization

3. (optional)Try DeepSpeed for accelerated inference: https://github.com/microsoft/DeepSpeed-MII/tree/main/examples/benchmark/txt2img )

4. Hyperparameter selection

5. Set up training log, potentially using wandb

6. Set up training validation(fix random seed, pick 5 images, visualize.不算fid了因为要花太长时间)7.Training:

Training UNet and encoder (These only teach the model one concept using 3-5 images:

<https://github.com/huggingface/diffusers/tree/main/examples/dreambooth>

<https://huggingface.co/blog/dreambooth>

Trainning from command line example on Huggingface: <https://huggingface.co/docs/diffusers/training/text_inversion>

7. Try other tricks when training is running (CollosalAI, change PIL to cv2.imread)

8.朋友圈采访真人评价

疑问:

VAE怎么控制output size的，是只改变latent code（sampled noise）的size吗？

文字最长只能77个token,但为什么encode出来length能是246，243？

ans：77 是 CLIP model 的 max length，它的 pos\_embbeding length，tokenizer能处理任意长度。

merge.txt还没搞懂

Training speed：

512\*512, 4\*100 images, 3000 max steps 一次训练要3小时

256\*256..(follows the above) 一次训练1小时

Tokenizer doc: <https://huggingface.co/docs/transformers/v4.24.0/en/main_classes/tokenizer#transformers.PreTrainedTokenizer>

Experiment notes:

1.Tokenizer遇到其他语言会随机encoder到词表中的一个字符。（see distillBart notebook)

2.Set random seed=42 for reproducibility.

3.Clip tokenizer **ignores space**

4.不管输入在不在vocab里都会decode回原来的

1.tokenizer.encode() will call **tokenizer.tokenize()** first

2**.tokenizer.convert\_tokens\_to\_ids()** will encode without calling tokenize() (and will thus not add the end of word token “<w/>”

5.DistillBart summerize的时候自动翻译成英文了。。。。

6.Kornia做augmentation比albumentation更快

7. Merges.txt的解释：训练的时候merge base vocab（characters）的记录

<https://github.com/huggingface/transformers/issues/4777>

<https://github.com/huggingface/transformers/issues/1083#issuecomment-524303077>

8. CLIP Tokenizer用的byte-level BPE,有空的时候研究下怎么encode各种语言字符的。

<https://huggingface.co/transformers/v4.6.0/_modules/transformers/models/clip/tokenization_clip.html>

<https://huggingface.co/docs/transformers/tokenizer_summary>

Citations: FID, clip